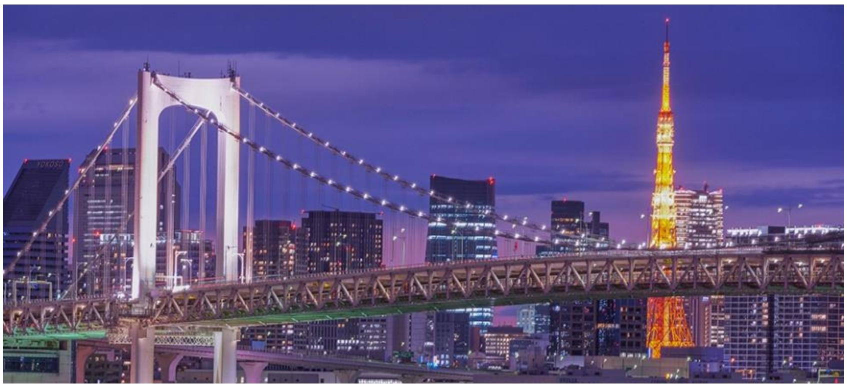
Research of Ramen restaurants in Tokyo

Norimoto Ichikawa

July 4, 2021

1. Introduction

There are various types of restaurants in Tokyo where has the most restaurants listed in the Michelin guide in the world. Especially Ramen (Japanese noodle) is a unique and a representative food of Japan which is strongly recommend for foreign tourists.



Its taste and style are completely unique in each restaurant, I strongly recommend trying several restaurants when foreigners visit to Tokyo. But there are a lot of Ramen restaurants anywhere in Tokyo. This research will show which location in Tokyo is convenient to stay and enjoy Ramens.

The central area of Tokyo is divided into 23 special wards (called "Ku"). This research only focuses on the 23 special wards area, and compares each ward.

2. Data Description

I used the following data source.

1) Wikipedia "Special wards of Tokyo"

https://en.wikipedia.org/wiki/Special_wards_of_Tokyo

I got each sward information from the table in Wikipedia scraping with BeautifulSoup.

2) Geocoder

I got Latitude and Longitude of each special ward using Geocoder.

3) Forsquare API

https://developer.foursquare.com/docs/api-reference/venues/explore/

I got Ramen restaurants in each sward using Forsquare API. Actually I had wanted to use popularity data of each restaurant such as rating and number of likes, however "Get Details of a Venue" API is Premium endpoint which only 50 calls per day is acceptable in my free account. Thus, I gave up to use popularity data, and used only location data in this research.

4) folium

I used folium to show maps with location data.

3. Methodology

3.1 Scraping Wikipedia using BeautifulSoup

Scraping "List of special wards" table from "Special wards of Tokyo" in Wikipedia using BeautifulSoup which is an HTML scraping tool.

https://en.wikipedia.org/wiki/Special_wards_of_Tokyo

List of special wards [edit]

No.	Flag	Name	Kanji	Population (as of October 2016)	Density (/km^2)	Area (km^2)	Major districts
01		Chiyoda	千代田区	59,441	5,100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi, Akihabara, Yūrakuchō, Iidabashi, Kanda
02		Chūō	中央区	147,620	14,460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōbori, Shinagawa, Tsukishima, Kachidoki, Tsukuda
03		Minato	港区	248,071	12,180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppongi, Toranomon, Aoyama, Azabu, Akasaka
04		Shinjuku	新宿区	339,211	18,620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ichigaya, Yotsuya, Sendagaya, Yoyogi

Since several tables are existing in the page, I selected only "wilitable sortable" class which only the target table uses. After scraping, I cleaned data such as removing rows of "No.", "Flag", and "Kanji", and renaming "Population (as of October 2016)" to "Population", and "OtaŌta" to "Ōta".

The result of scraping was as below,

	Name	Population	Density(/km2)	Area(km2)	Major districts
0	Chiyoda	0059,441	05,100	011.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	Chūō	0147,620	14,460	010.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	Minato	0248,071	12,180	020.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	Shinjuku	0339,211	18,620	018.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	Bunkyō	0223,389	19,790	011.29	Hongō, Yayoi, Hakusan
5	Taitō	0200,486	19,830	010.11	Ueno, Asakusa
6	Sumida	0260,358	18,910	013.77	Kinshichō, Morishita, Ryōgoku
7	Kōtō	0502,579	12,510	040.16	Kiba, Ariake, Kameido, Tōyōchō, Monzennakachō,...
8	Shinagawa	0392,492	17,180	022.84	Shinagawa, Gotanda, Ōsaki, Hatanodai, Ōimachi,...
9	Meguro	0280,283	19,110	014.67	Meguro, Nakameguro, Jiyugaoka, Komaba, Aobadai
10	Ōta	0722,608	11,910	060.66	Ōmori, Kamata, Haneda, Den-en-chōfu
11	Setagaya	0910,868	15,690	058.05	Shimokitazawa, Kinuta, Karasuyama, Tamagawa
12	Shibuya	0227,850	15,080	015.11	Shibuya, Ebisu, Harajuku, Daikanyama, Hiroo
13	Nakano	0332,902	21,350	015.59	Nakano
14	Suginami	0570,483	16,750	034.06	Kōenji, Asagaya, Ogikubo
15	Toshima	0294,673	22,650	013.01	Ikebukuro, Komagome, Senkawa, Sugamo
16	Kita	0345,063	16,740	020.61	Akabane, Ōji, Tabata
17	Arakawa	0213,648	21,030	010.16	Arakawa, Machiya, Nippori, Minamisenju
18	Itabashi	0569,225	17,670	032.22	Itabashi, Takashimadaira
19	Nerima	0726,748	15,120	048.08	Nerima, Ōizumi, Hikarigaoka
20	Adachi	0674,067	12,660	053.25	Ayase, Kitasenju, Takenotsuka
21	Katsushika	0447,140	12,850	034.80	Tateishi, Aoto, Kameari, Shibamata
22	Edogawa	0685,899	13,750	049.90	Kasai, Koiwa

3.2 Get Latitude and Longitude using Geocoder

Searching address such as “Chiyoda-ku, Tokyo” and getting Latitude and Longitude of typical location of each ward. Latitude and Longitude are added to the wards list as below,

	Name	Population	Density(/km2)	Area(km2)	Major districts	Latitude	Longitude
0	Chiyoda	0059,441	05,100	011.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.693930	139.753711
1	Chūō	0147,620	14,460	010.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	35.670572	139.771988
2	Minato	0248,071	12,180	020.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.658017	139.751546
3	Shinjuku	0339,211	18,620	018.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	35.693798	139.703440
4	Bunkyō	0223,389	19,790	011.29	Hongō, Yayoi, Hakusan	35.707595	139.752210
5	Taitō	0200,486	19,830	010.11	Ueno, Asakusa	35.712595	139.779986
6	Sumida	0260,358	18,910	013.77	Kinshichō, Morishita, Ryōgoku	35.710707	139.801540
7	Kōtō	0502,579	12,510	040.16	Kiba, Ariake, Kameido, Tōyōchō, Monzennakachō,...	35.672823	139.817290
8	Shinagawa	0392,492	17,180	022.84	Shinagawa, Gotanda, Ōsaki, Hatanodai, Ōimachi,...	35.609160	139.730161
9	Meguro	0280,283	19,110	014.67	Meguro, Nakameguro, Jiyugaoka, Komaba, Aobadai	35.641490	139.698273
10	Ōta	0722,608	11,910	060.66	Ōmori, Kamata, Haneda, Den-en-chōfu	35.561361	139.716081
11	Setagaya	0910,868	15,690	058.05	Shimokitazawa, Kinuta, Karasuyama, Tamagawa	35.646544	139.653222
12	Shibuya	0227,850	15,080	015.11	Shibuya, Ebisu, Harajuku, Daikanyama, Hiroo	35.663687	139.697791
13	Nakano	0332,902	21,350	015.59	Nakano	35.707342	139.663801
14	Suginami	0570,483	16,750	034.06	Kōenji, Asagaya, Ogikubo	35.699731	139.636138
15	Toshima	0294,673	22,650	013.01	Ikebukuro, Komagome, Senkawa, Sugamo	35.726128	139.716693
16	Kita	0345,063	16,740	020.61	Akabane, Ōji, Tabata	35.752839	139.733519
17	Arakawa	0213,648	21,030	010.16	Arakawa, Machiya, Nippori, Minamisenju	35.736093	139.783403
18	Itabashi	0569,225	17,670	032.22	Itabashi, Takashimadaira	35.751074	139.709194
19	Nerima	0726,748	15,120	048.08	Nerima, Ōizumi, Hikarigaoka	35.735700	139.651607
20	Adachi	0674,067	12,660	053.25	Ayase, Kitasenju, Takenotsuka	35.774811	139.804537
21	Katsushika	0447,140	12,850	034.80	Tateishi, Aoto, Kameari, Shibamata	35.743454	139.847229
22	Edogawa	0685,899	13,750	049.90	Kasai, Koiwa	35.707014	139.868367

Plotting each sward location using folium is as below,

3.3 Get Ramen restaurants in each sward using Foursquare

"Get Venue Recommendations" API of Foursquare is used to get Ramen restaurants in each sward.

https://developer.foursquare.com/docs/api-reference/venues/explore/

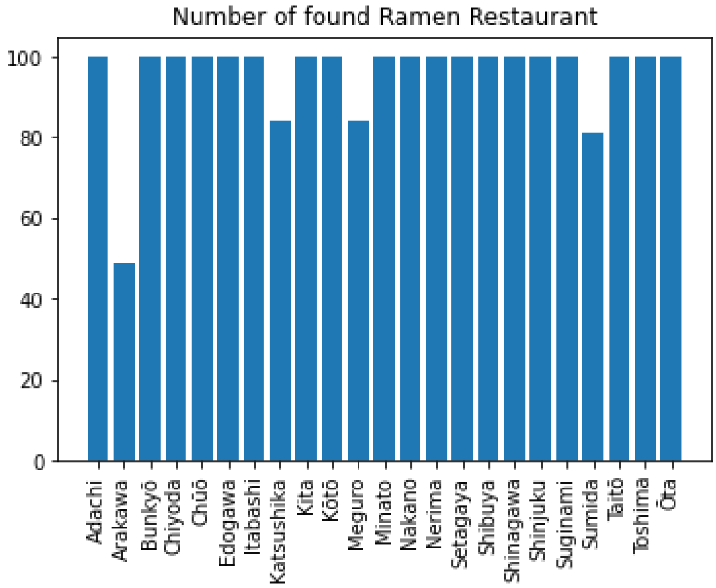
To select Ramen restaurant only, categoryid was set as "55a59bace4b013909087cb24" which is "Ramen Restaurant".

https://developer.foursquare.com/docs/build-with-foursquare/categories/

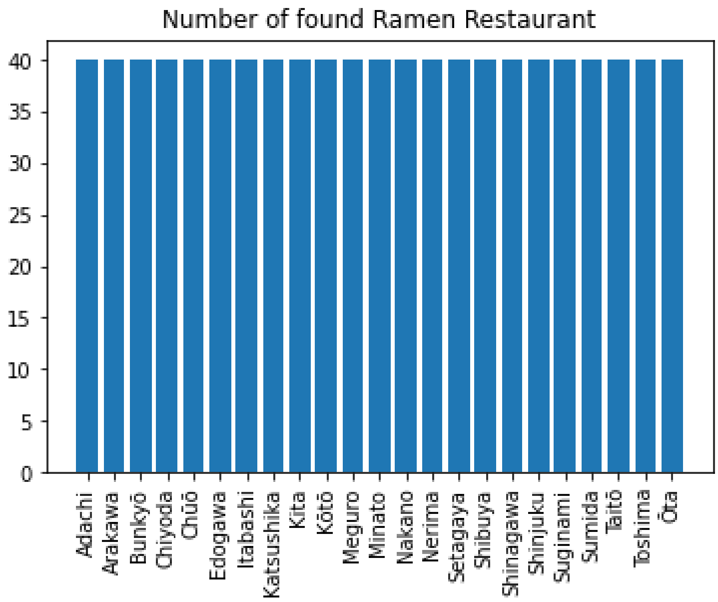
Radius is set as 100,000m because restaurants are searched by sward name and radius is not used.

sortByPopularity = 1 option was used to get only higher popularity restaurants were used.

At first, the number of restaurants in each sward had been limited as 100 to avoid over API calls, however I found that some swards didn't achieve to 100.



To align number of restaurants in each sward, I limited to 40.



Total 920 (40 x 23 swards) restaurants were found as below,

	Ward	Ward Latitude	Ward Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chiyoda	35.693930	139.753711	Tsujita (つじ田)	35.698808	139.770099	Ramen Restaurant
1	Chiyoda	35.693930	139.753711	雷 東京本丸店	35.681882	139.766340	Ramen Restaurant
2	Chiyoda	35.693930	139.753711	Tokyo Ramen Street (東京ラーメンストリート)	35.680142	139.767873	Ramen Restaurant
3	Chiyoda	35.693930	139.753711	Honda (麺処 ほん田)	35.698017	139.774240	Ramen Restaurant
4	Chiyoda	35.693930	139.753711	Bushoya (家系らーめん 武将家)	35.700019	139.771998	Ramen Restaurant
...	...	...	...	...	...	...	...
915	Edogawa	35.707014	139.868367	麺屋こころ 小岩店	35.735485	139.881764	Ramen Restaurant
916	Edogawa	35.707014	139.868367	Ramen Yoshibe (ラーメン ヨシベー)	35.664727	139.860224	Ramen Restaurant
917	Edogawa	35.707014	139.868367	Misoichi (味噌一)	35.690402	139.882639	Ramen Restaurant
918	Edogawa	35.707014	139.868367	宝来軒	35.669893	139.874578	Ramen Restaurant
919	Edogawa	35.707014	139.868367	Chibaki-Ya (ちばき屋)	35.662332	139.875219	Ramen Restaurant

920 rows × 7 columns

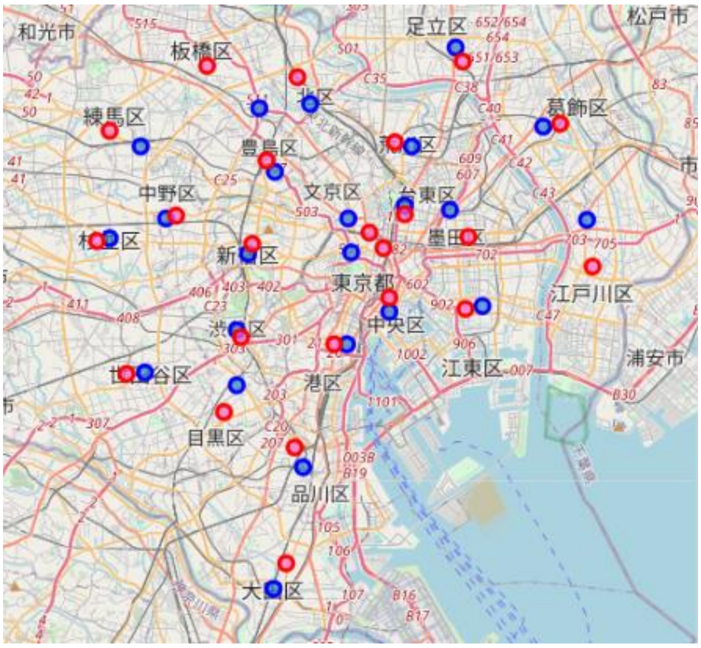
For instance, Ramen restaurants in Minato-ku are as below,

3.4 Calculate centroid of Ramen restaurant in each sward

To see deviation of distance from centroid in each sward, centroids of all Ramen restaurant in each sward was calculated using “groupby().mean()” method.

Ward	Ward Latitude	Ward Longitude	Venue Latitude	Venue Longitude
Adachi	35.774811	139.804537	35.769489	139.808063
Arakawa	35.736093	139.783403	35.737450	139.775122
Bunkyō	35.707595	139.752210	35.702008	139.762715
Chiyoda	35.693930	139.753711	35.695866	139.769709
Chūō	35.670572	139.771988	35.676320	139.772320
Edogawa	35.707014	139.868367	35.688505	139.870983
Itabashi	35.751074	139.709194	35.767604	139.683700
Katsushika	35.743454	139.847229	35.744980	139.854937
Kita	35.752839	139.733519	35.763150	139.727331
Kōtō	35.672823	139.817290	35.672059	139.809386
Meguro	35.641490	139.698273	35.630854	139.692023
Minato	35.658017	139.751546	35.657773	139.745247
Nakano	35.707342	139.663801	35.708450	139.668845
Nerima	35.735700	139.651607	35.742142	139.636152
Setagaya	35.646544	139.653222	35.646086	139.644433
Shibuya	35.663687	139.697791	35.660598	139.700023
Shinagawa	35.609160	139.730161	35.616844	139.726483
Shinjuku	35.693798	139.703440	35.697620	139.705293
Suginami	35.699731	139.636138	35.698580	139.630326
Sumida	35.710707	139.801540	35.700120	139.810738
Taitō	35.712595	139.779986	35.709333	139.779747
Toshima	35.726128	139.716693	35.730374	139.712316
Ōta	35.561361	139.716081	35.571402	139.721927

Blue is the typical location of each sward, and red is the centroid of Ramen restaurants in each sward.

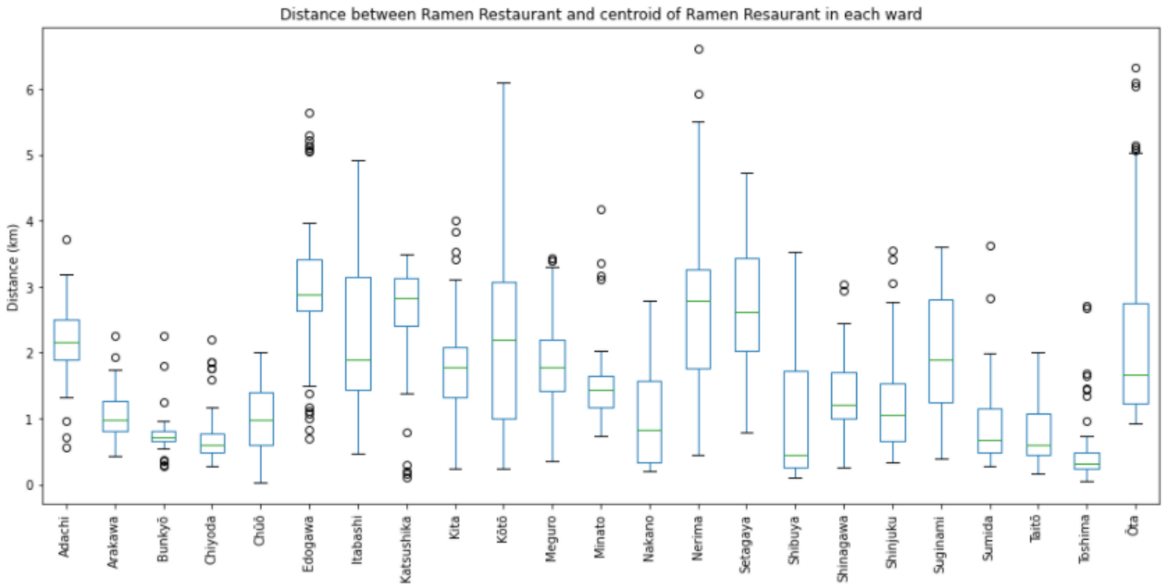


3.5 Calculate distance between each Ramen restaurant and a centroid

To calculate a distance between 2 locations in Latitude and Longitude, geopy.distance was used. Distances of each restaurant are as below,

Adachi	Arakawa	Bunkyō	Chiyoda	Chūō	Edogawa	Itabashi	Katsushika	Kita	Kōtō	...	Nerima	Setagaya	Shibuya	Shinagawa	Shinjuku	Suginami	Sumida	Taitō	Toshi
3.715513	0.740585	0.725355	0.328345	0.476824	1.107217	3.338382	2.909538	1.687451	3.654306	...	2.352257	2.670098	0.416073	1.018255	1.447752	0.664403	0.623493	0.617763	0.288
1.700491	0.908688	0.712774	1.581240	0.575735	5.215615	0.786634	3.019877	2.465635	1.259372	...	2.845037	3.768844	0.296912	3.043850	0.344965	1.033526	0.665213	0.427579	0.277
2.133127	1.143981	0.691872	1.752502	0.367481	0.992028	2.265685	2.529932	1.307569	2.413759	...	1.549708	2.264397	2.107132	1.709780	1.693813	2.820245	0.296310	0.743831	1.428
2.257037	0.840990	0.787535	0.474461	1.823483	5.057924	2.187555	2.519922	2.225116	4.017032	...	2.301812	2.365480	1.422315	0.247790	1.082816	3.266742	0.936978	0.589185	0.227
2.808380	0.722718	1.798864	0.505175	0.734498	2.765419	0.461164	2.970562	0.300712	3.118486	...	1.610847	3.849582	1.985996	1.692662	0.651672	2.815684	0.399740	0.161376	0.741
2.218167	0.846425	0.775960	0.448851	1.948922	5.110608	2.332456	2.888723	1.776672	0.979709	...	3.835448	4.123214	0.183586	0.695105	3.057306	2.771145	1.155246	0.384023	0.250
3.149593	1.578608	0.694175	0.720856	1.085234	2.968077	0.978494	3.030870	1.629006	2.869472	...	1.737374	2.644488	0.502001	1.209750	0.678532	1.840118	1.171640	0.534164	0.400
2.649948	1.683418	0.826926	0.549560	1.399624	2.759933	1.210279	2.752558	1.426782	3.351639	...	1.611481	4.638400	0.447994	1.032861	0.770620	2.200889	0.413963	0.769467	0.438
2.329732	0.721189	2.255764	1.763365	0.574853	3.432888	1.668724	3.146317	3.842303	2.351832	...	3.262921	2.668218	0.252054	1.000309	0.605121	1.942732	0.634557	0.529198	0.246
2.204741	0.434172	0.625533	0.350186	1.504082	2.786925	2.230152	3.273368	0.499379	3.298753	...	4.493330	2.442707	0.326620	1.403538	0.376647	2.681052	0.337669	2.012570	0.145
2.003168	2.247907	0.726566	1.852983	0.500323	5.647414	3.186393	3.050517	1.952702	6.073305	...	1.826537	1.931530	1.991138	0.839559	0.917519	1.815080	0.515087	1.449428	0.246
1.799421	0.820630	0.674874	0.633589	1.599562	2.430891	4.913928	1.375392	1.491949	3.296493	...	5.935714	0.977253	0.430854	1.356751	0.545938	3.583108	0.680374	0.891598	0.315
2.062605	1.061269	0.537415	0.565441	0.384835	0.690393	1.943147	3.304778	1.229792	2.373479	...	0.449827	2.701604	0.118069	2.173651	0.510663	1.362315	0.434172	0.436042	0.303
1.803030	1.742871	0.670176	0.723036	1.389563	5.156277	4.737756	2.983303	1.758198	1.447066	...	3.263408	1.341314	2.656720	1.139414	1.604375	2.241467	1.405090	0.455375	0.202
2.217136	0.947877	0.752020	0.482536	0.689643	5.071447	1.841779	2.443566	0.313908	0.244274	...	5.284027	4.151850	1.751054	1.153899	0.851269	1.016024	1.388471	0.792147	0.401
2.499248	0.877787	0.736844	0.292928	0.529233	1.166833	1.179498	2.299241	1.823420	1.427269	...	2.360312	3.370513	2.981476	0.930328	1.427729	1.924656	0.542018	1.084820	0.310
1.967533	0.715386	0.718640	0.565885	1.176718	3.145136	4.704592	1.804311	1.325060	3.141467	...	2.678816	0.819667	0.235166	2.178375	1.056061	2.836468	0.405994	1.721596	0.325
2.372733	0.905289	0.615524	0.706791	1.140938	2.872430	3.220429	2.644943	1.844106	2.772956	...	1.844370	2.719373	1.830802	2.152130	0.469424	1.349819	0.585079	0.758968	0.401
1.973100	0.755345	0.908744	0.599422	0.835731	2.886021	1.098600	2.599223	1.325165	2.494714	...	5.501506	2.235287	0.518456	1.511739	1.245924	0.734803	1.091094	0.415274	0.282

The smaller distribution and the shorter average of distance may be the more convenient to visit several restaurants in the location. The box plot of the distribution of distances in each ward is as below,



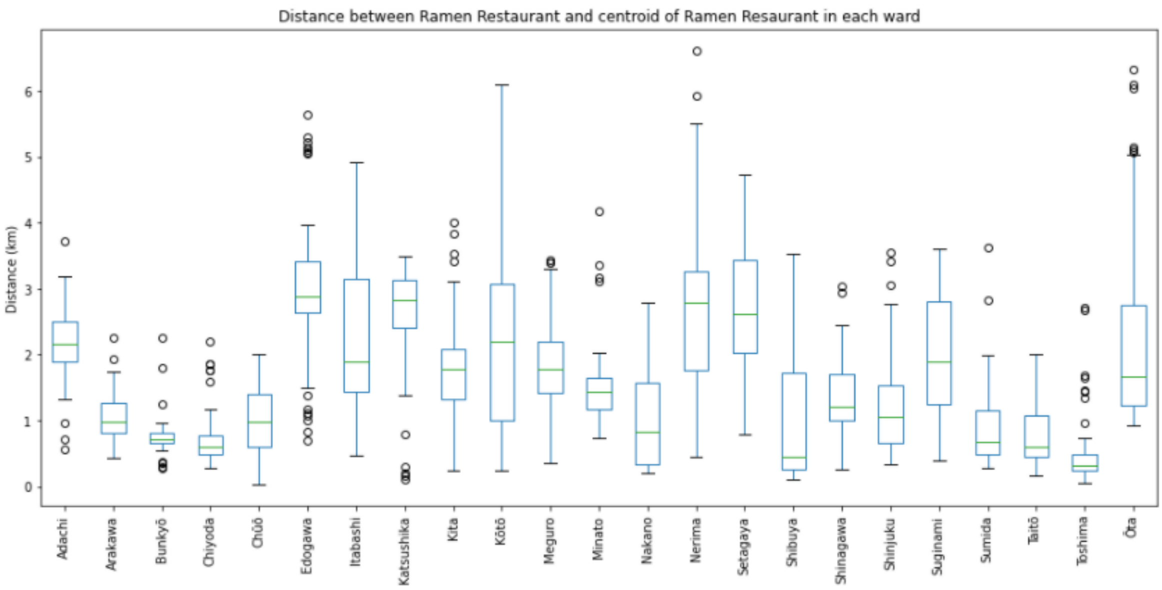
3.6 Clustering based on distribution of distance

The distribution of distance of each restaurant was summarized using describe() method and clustered using K-mean (K=5).

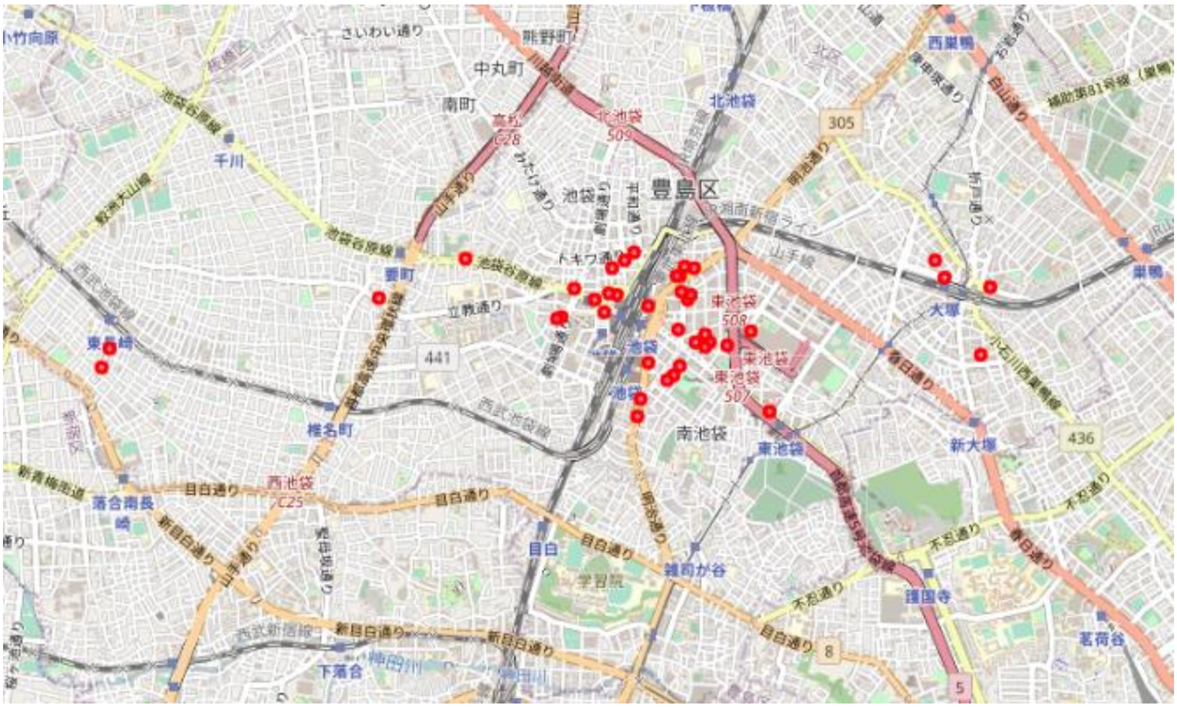
Ward	mean	std	min	25%	50%	75%	max	Cluster
Adachi	2.192385	0.655676	0.566281	1.898154	2.168047	2.506668	3.715513	3
Arakawa	1.067880	0.405522	0.434172	0.804309	0.972008	1.256572	2.247907	0
Bunkyō	0.775298	0.345828	0.277652	0.647688	0.721997	0.809847	2.255764	0
Chiyoda	0.759070	0.497580	0.272763	0.480517	0.606993	0.772592	2.207177	0
Chūō	1.030066	0.510369	0.033412	0.603748	0.972691	1.392079	2.000742	0
Edogawa	3.012988	1.326232	0.690393	2.633539	2.879225	3.424366	5.647414	1
Itabashi	2.307524	1.268142	0.461164	1.437933	1.896246	3.152462	4.913928	1
Katsushika	2.492202	0.952937	0.103615	2.407485	2.820641	3.126736	3.487477	1
Kita	1.748695	0.937671	0.230776	1.320687	1.771693	2.081480	4.010110	3
Kōtō	2.161444	1.419227	0.244274	1.007691	2.196406	3.066265	6.100544	4
Meguro	1.887963	0.842603	0.356947	1.408465	1.775668	2.203347	3.427148	3
Minato	1.565996	0.721836	0.740683	1.166174	1.429586	1.645702	4.175463	3
Nakano	1.007666	0.739849	0.206006	0.328399	0.828222	1.560954	2.778450	2
Nerima	2.811174	1.418058	0.449827	1.766016	2.780968	3.271596	6.616007	4
Setagaya	2.595546	1.088143	0.787674	2.019945	2.614474	3.436831	4.727531	1
Shibuya	0.920224	0.934387	0.112243	0.266315	0.440627	1.722212	3.529886	2
Shinagawa	1.404124	0.616891	0.247790	0.999770	1.212159	1.702775	3.043850	2
Shinjuku	1.212221	0.812729	0.334750	0.651284	1.062309	1.529580	3.545850	2
Suginami	2.006736	0.931981	0.388779	1.244425	1.900630	2.799226	3.598740	3
Sumida	0.924150	0.693673	0.272807	0.491162	0.672794	1.159345	3.621577	2
Taitō	0.769662	0.438125	0.161376	0.455262	0.592657	1.078597	2.012570	0
Toshima	0.596362	0.646392	0.057170	0.246848	0.312830	0.487332	2.717488	0
Ōta	2.396526	1.668820	0.917471	1.218075	1.660156	2.752502	6.330553	4

4. Results

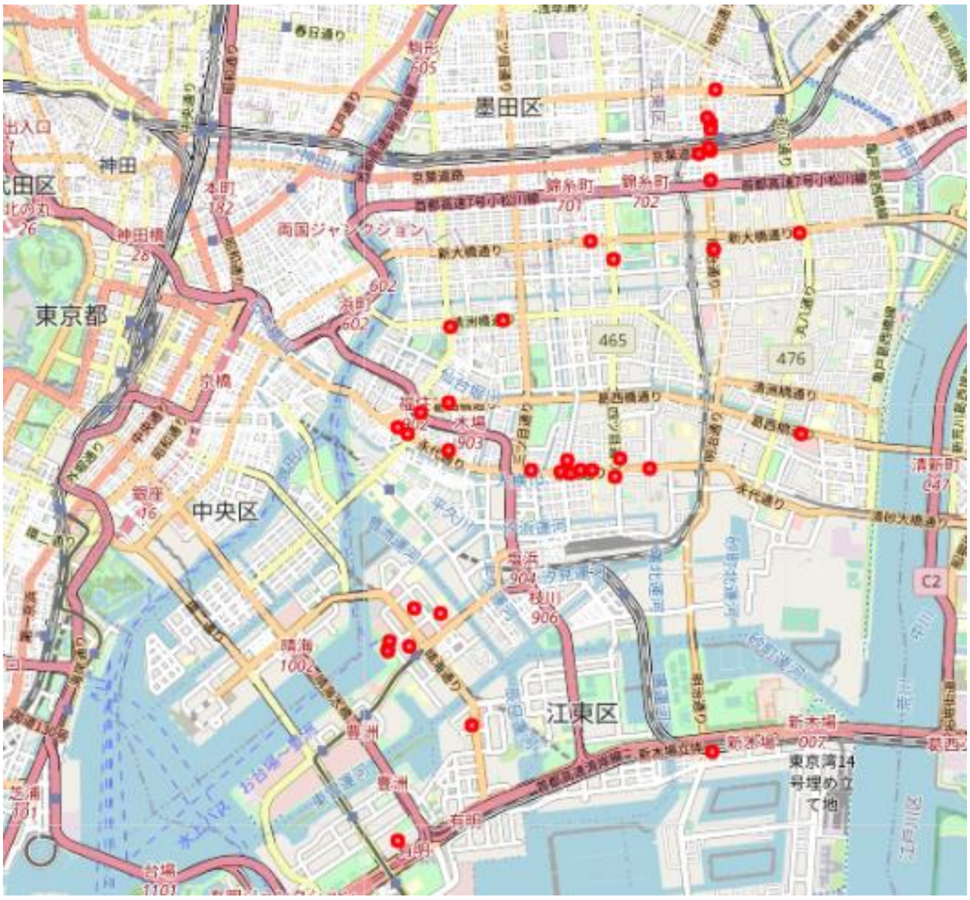
Looking on the distribution of the box plot, Toshima-ku is small distribution and short average.



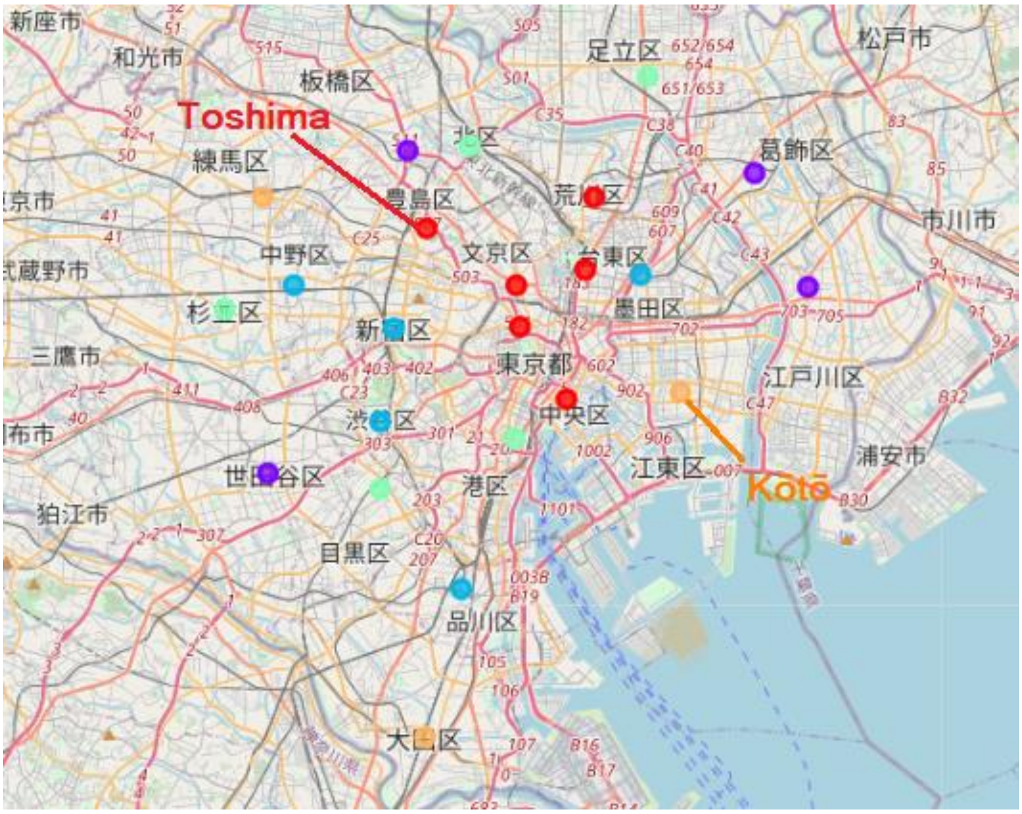
This is the map of each Ramen restaurant in Toshima-ku. There is a large terminal station called Ikebukuro station in Toshima-ku. Many Ramen restaurants are gathering around the station and looks convenient to visit many restaurants.



Contrary, Kōtō-ku is the largest distribution. It looks hard to visit to many restaurants.



This is the clustering result using k-means (k=5). Kōtō-ku is orange sward. It looks orange swards are a little far from central area of Tokyo. Red swards which Toshima-ku belongs locate in the central area of Tokyo and look convenient to visit Ramen restaurants and other sightseeing spots.



5. Discussion

As I mentioned in section 2, I had supposed to use popularity data of each restaurant. Because taste of popular restaurant and non-populer restaurant are completely difference in case of Ramen restaurant, however I could not use it due to restriction of API call. But I think, interesting result was made even though only location data was used.

Although I am commuting to our office in Tokyo every day, I didn’t know which area is the most area where many Ramen restaurant are existing. Looking data can make it clear for me.

I got another curiosity during the research that difference of characteristics depending on food type which means how other foods are different from Ramen restaurant. I will try to study later.

6. Conclusion

Now I have a confidence that I can recommend some areas in Tokyo to foreign tourists and Ramen lovers.

I can use this data science methodology to other geographical projects.